

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-146760

(43)Date of publication of application : 06.06.1995

(51)Int.Cl.

G06F 3/06

G06F 3/06

(21)Application number : 05-314487

(71)Applicant : MUTOH IND LTD

(22)Date of filing : 19.11.1993

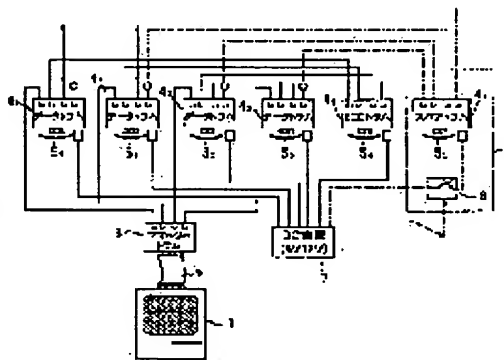
(72)Inventor : UMITAKE TAKAFUMI

(54) DISK ARRAY DEVICE

(57)Abstract:

PURPOSE: To provide a disk array device by which a data recovery processing is automatically performed without interposing an operator.

CONSTITUTION: A spare disk device 55 is preliminarily provided in addition to disk devices 50 to 53 dispersedly storing data and a disk device 54 storing parity information ECC. If the data destruction in either one of the disk devices 50 to 53 is detected, the data of the disk device where data is destroyed is automatically restored from the data of other disk device based on the error detection result and the data is written in the spare disk 55.



LEGAL STATUS

[Date of request for examination]

27.11.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3009987

[Date of registration]

03.12.1999

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-146760

(43) 公開日 平成7年(1995)6月6日

(51) Int.Cl.⁶

G 0 6 F 3/06

識別記号

5 4 0

庁内整理番号

3 0 5 Z

F I

技術表示箇所

審査請求 未請求 請求項の数 3 F D (全 5 頁)

(21) 出願番号 特願平5-314487

(22) 出願日 平成5年(1993)11月19日

(71) 出願人 000238566

武藤工業株式会社

東京都世田谷区池尻3丁目1番3号

(72) 発明者 海嶽 尚文

東京都世田谷区池尻3丁目1番3号 武藤

工業株式会社内

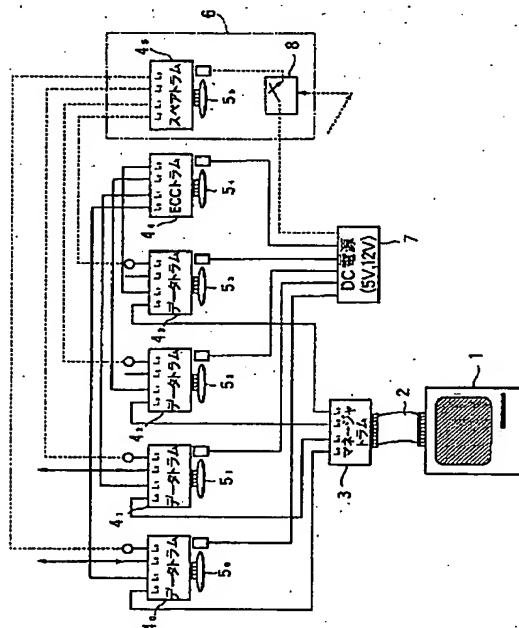
(74) 代理人 弁理士 伊丹 勝

(54) 【発明の名称】 ディスクアレイ装置

(57) 【要約】

【目的】 オペレータの介在なしに自動的にデータ復帰処理がなされるようにしたディスクアレイ装置を提供する。

【構成】 データを分散記憶するディスク装置 50 ～ 53 及びパリティ情報 ECC を記憶するディスク装置 54 の他に、スペアディスク装置 55 が予め設けられている。ディスク装置 50 ～ 53 のいずれかでのデータ破壊が検出されたら、そのエラー検出結果に基づいてデータが破壊がされたディスク装置のデータが他のディスク装置のデータから自動復元されて、スペアディスク装置 55 に書き込まれる。



【特許請求の範囲】

【請求項1】 データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置において、

前記複数台のディスク装置とは別個に設けられた予備ディスク装置と、

前記複数台のディスク装置でのデータ破壊を検出するエラー検出手段と、

このエラー検出手段の出力に基づいてデータ破壊が検出されたディスク装置のデータを他のディスク装置のデータから復元して前記予備ディスク装置に格納するデータ復元処理手段とを有することを特徴とするディスクアレイ装置。

【請求項2】 前記データ復元処理手段は、通常動作の合間にデータの復元処理を実行することを特徴とする請求項1に記載のディスクアレイ装置。

【請求項3】 前記データ復元処理手段は、データ復元処理時間間隔又はデータ復元処理1回の復元情報量の大きさ単位の少なくとも一方を可変として、通常動作の合間にデータの復元処理を実行する機能を有することを特徴とする請求項1に記載のディスクアレイ装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置に関する。

【0002】

【従来の技術】 大量のデータを複数台の小型ディスク装置に分散して記憶することにより危険分散を図るディスクアレイ装置として、RAID (Redundant Arrays of Inexpensive Disks) が知られている。RAIDには、1から5までのレベルがある。2～4のレベルでは、複数のディスク装置がデータを分散記憶するものと、エラーチェック用のパリティ情報を記憶するものとに用途分けされている。レベル5では全てのディスク装置にデータと共にパリティ情報も分散して記憶される。このようなRAID装置の中のいずれかのディスクでデータが破壊されたとき、従来は故障ディスクに代わって新しいディスクを装填し、データ復帰ボタン等により復帰プログラムをスタートさせ、破壊されたデータ分を新しいディスクに復帰させる方式が採られている。

【0003】

【発明が解決しようとする課題】 従来のRAIDでのデータ復帰方式は、オペレータにとって新しいディスクの装填とデータ復帰という操作を必要とし、操作ミスが生じ易い。また、このデータ復帰処理の間、ディスク装置のアクセスは中止しなければならない。本発明は、オペレータの介在なしに自動的にデータ復帰処理がなされるようにしたディスクアレイ装置を提供することを目的としている。

【0004】

【課題を解決するための手段】 本発明は、データを分散記憶する複数台のディスク装置を配置して構成されるディスクアレイ装置において、前記複数台のディスク装置でのデータ破壊を検出するエラー検出手段と、このエラー検出手段の出力に基づいてデータ破壊が検出されたディスク装置のデータを他のディスク装置のデータから自動復元して格納するための予備のディスク装置とを有することを特徴としている。本発明において好ましくは、データ復元処理手段は、通常動作の合間にデータの復元処理を実行する。この場合、効率的な復元処理を行うためには、データ復元処理の時間間隔を可変とする機能、又は1回のデータ復元処理で復元すべき情報量の大きさ単位を可変とする機能の少なくとも一方をオプション機能としてシステムに組み込むことが有効である。

【0005】

【作用】 本発明によれば、ディスクアレイ装置には予め予備のディスク装置が設けられていて、あるディスクにデータエラーがあった場合には、自動的にその故障ディスクのデータが他のディスクのデータによって復元されて予備のディスク装置に格納される。従って、大切なデータがオペレータの操作ミスで復帰できなくなるといった事態が防止できる。特に本発明を、並列データ転送可能な通信プロセッサ（以下、トラムという）をツリー状に配置して複数のディスク装置の並列処理を可能としたシステムに適用すれば、トラム間通信によってデータエラー検出後直ちに、通常動作と並列にデータ復元処理ができる。また、ディスク装置とホストコンピュータとのデータ授受の合間を利用してデータ復元処理を行うことにより、通常動作に何等影響を与えることなくデータの復元が可能になる。

【0006】

【実施例】 以下、図面を参照して、本発明の実施例を説明する。図1は、本発明の一実施例に係るディスクアレイ装置のブロック構成を示す。この装置は、5台のディスク装置50～54を有する。これらのディスク装置50～54のうち4台のディスク装置50～53がデータを分散記憶するものであり、残り1台のディスク装置54はパリティ情報ECCを記憶するものである。これらのディスク装置50～54に1対1に対応してアクセス手段としてのデータトラム40～43及びECCトラム44が設けられている。これらのトラム40～44は例えば、並列データ転送が可能な4つのシリアルポートを持つ通信プロセッサであり、好適にはインモス社のトラムコンピュータ（商標）等が使用される。またこれらのトラム40～44にはそれぞれSCSIポート（パラレル）が設けられており、このSCSIポートを介してそれぞれディスク装置50～54に接続されている。

【0007】 ホストコンピュータ1をアクセスするデータ転送手段として、マネージャトラム3が設けられてい

る。このマネージャトラム3もデータトラム40~43及びECCトラム44と同様の4つのシリアルポートを持つ通信プロセッサである。マネージャトラム3のバスポートは、SCSIインタフェース2を介してホストコンピュータ1と接続されている。マネージャトラム3の4つのシリアルポートは、データトラム40~43の各1つのシリアルポートと接続されている。ECCトラム44の4つのシリアルポートは、データトラム40~43の各1つのシリアルポートと接続されている。

【0008】以上の基本構成の他、RAID装置内には、ディスク装置に故障があった場合の予備装置6として、スペアディスク装置55及びスペアトラム45が設けられている。スペアトラム45も、データトラム40~43及びECCトラム44と同様の4つのシリアルポートを持つ通信プロセッサであり、その4つのシリアルポートは破線で示すようにそれぞれデータトラム40~43の各1つのシリアルポートに接続されている。この予備装置6の電源スイッチ8は、RAID装置が正常動作している限りオフであり、通常直流電源7の出力は基本構成部のみに供給される。ディスク装置に故障があった場合には、マネージャトラム3の制御により、又はエラーメッセージに従ったオペレータの手動操作より電源スイッチ8がオンされる。

【0009】この実施例のRAID装置において、記憶すべきデータは4つのディスク装置50~53にビット単位又はセクタ単位に分割されて書き込まれる。残りのディスク装置54には、これらの分割書き込まれたデータのバリディ情報ECCが書き込まれる。バリディ情報ECCは例えば、マネージャトラム3で計算される。このデータの書き込み/読出しの詳細は省略する。

【0010】この実施例のRAID装置において、ディスク装置50~53のいずれかでのデータ破壊はSCSIインタフェース信号により検出することができる。データ破壊が検出されたら、残りのディスク装置のデータに基づいて破壊されたデータの内容を自動的に復元してこれをスペアディスク装置55に格納するという処理がなされる。この処理は例えば、マネージャトラム3のデータ復元プログラムによってなされる。あるいはスペアトラム45でこの処理を行うこともできる。

【0011】データ復元処理の流れは、簡単に示せば図2の通りである。SCSIインタフェース信号からディスク装置のエラー検出がなされる。エラーが検出されたら、この検出結果によりエラーステータスをオンにする(S1)。そして破壊されたディスク装置の電源スイッチをオフ(S2)、スペアディスク装置55の電源スイッチ8をオン(S3)して、データ復元処理を行う(S4)。破壊されたデータが全て復旧されたら、エラーステータスをオフにし(S5)、以後通常動作に戻る。

【0012】破壊されたディスクの内容は前述のように他のディスクから逆計算されるが、そのデータ復元処理

のアルゴリズムは次の通りである。4つのディスク装置50~53のデータをそれぞれA、B、C、Dで表すと、バリディ情報ECCは、次の数1で示される。なお以下の式で、“+”は排他的論理和(XOR)を表す。

【0013】

【数1】 $A+B+C+D=ECC$

【0014】データBが破壊されたと仮定して、これを残りのデータA、C、D、ECCから求める場合を説明すれば、数1の両辺にECC+BをXORすると、

【0015】

【数2】

$A+B+C+D+ECC+B=ECC+ECC+B$

【0016】となる。数2は、符号理論から、 $B+B=0$ 、 $ECC+ECC=0$ であるから、結局次の数3と等価になる。

【0017】

【数3】 $A+C+D+ECC=B$

【0018】従って数3に基づいて、破壊されたデータBが逆計算されて求められることになる。

【0019】この実施例において、破壊されたデータの復元処理は、例えば図3に示すようにホストコンピュータ1とのデータ授受の合間を利用して1セクタデータずつ行うことができる。この動作において、破壊されたデータの全てが復旧されたとき、エラーステータスをオフにしてエラーディスクのない通常動作を行う。これにより通常のデータ読み書き動作に何等影響を与えることなく、データ復元処理を実行することができる。また、各トラム3、40~45に並列動作可能な通信プロセッサを用いれば、データ修復をトランザクションと並列に実行することが可能である。

【0020】通常動作の合間を利用したデータ復元処理の場合、復元処理中にもトランザクション要求が発生したとすると、トランザクションが優先される。従ってデータ復元処理中に頻繁にトランザクションが行われると、復元処理完了までに多くの時間がかかる。この問題を解決するには、次の二つの機能のいずれかをオプション機能としてシステムに組み込むことが有効である。

①データ復元処理の時間間隔を可変とする機能

②1回のデータ復元処理で復元すべき情報量の大きさ単位を可変とする機能

【0021】①の方法は、復元処理の時間間隔Tを、図4(a)に例示したT1あるいはT2のように可変とするものであり、②の方法は、1回の復元処理で復元すべき情報量の大きさの単位Rを、図4(b)に例示したR1あるいはR2のように可変とする。これらの具体的な機能実現方法としては、例えばディップスイッチで選択してその内容をシステムソフトが読むという方法を用いることができる。

【0022】本発明は上記実施例に限られない。実施例ではバリディ情報ECCをデータと別に固定のディスク

5

に記憶するレベル2～4のRAIDを説明したが、ECCをデータと同様に分散記憶するレベル5にも同様に本発明を適用することができる。

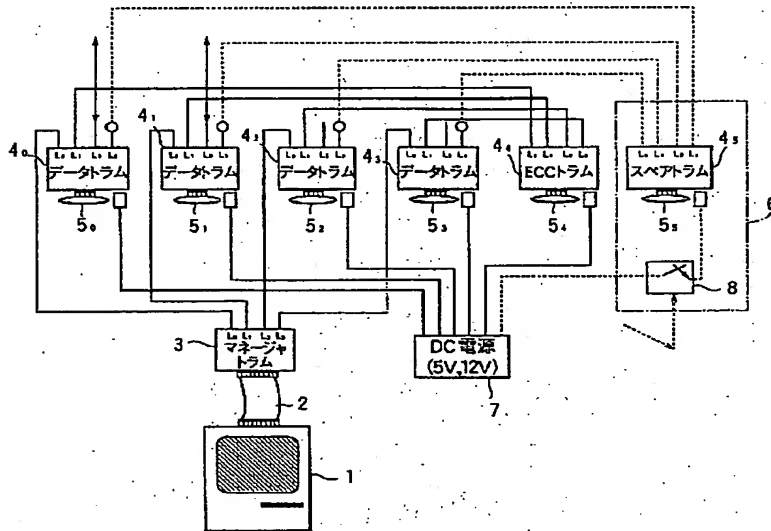
【0023】

【発明の効果】以上述べたように本発明によれば、予備のディスク装置を備えることで、データ破壊があった場合にオペレータの介入なしに自動的にデータ復帰処理がなされるようにして、ディスクアレイ装置の操作ミスを防止することができる。

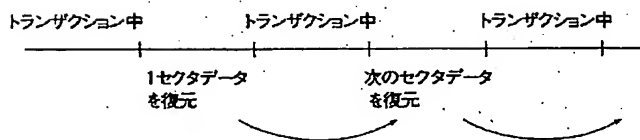
【図面の簡単な説明】

【図1】 本発明の一実施例に係るディスクアレイ装置を示す。

【図1】



【図3】



6

【図2】 同実施例のデータ復元処理の流れを示す。

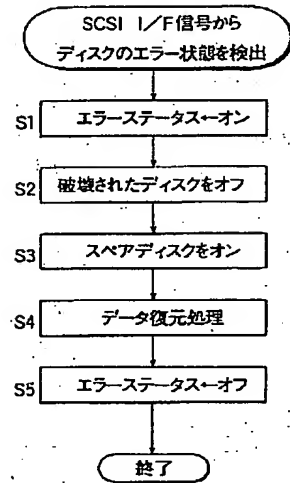
【図3】 同実施例のデータ復元の並列処理の様子を示す。

【図4】 他の実施例のデータ復元の並列処理の様子を示す。

【符号の説明】

1…ホストコンピュータ、2…SCSIインタフェース、3…マネージャトラム、40～43…データトラム、44…ECCトラム、50～54…ディスク装置、45…スベアトラム、55…スベアディスク装置、7…直流電源、8…電源スイッチ。

【図2】

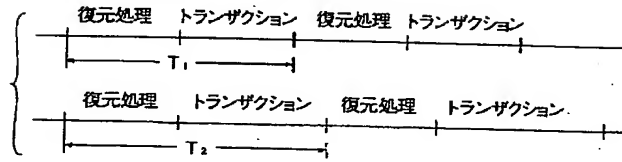


(5)

特開平7-146760

【図4】

(a) 復元処理時間間隔可変



(b) 復元処理単位情報量可変

